# Response to February 2025 Review

Sean Youn

February 2025

This document addresses comments from reviewers for the article "Positive Matrix Factorization of Large Real-Time Atmospheric Mass Spectrometry Datasets Using Error-Weighted Randomized Hierarchical Alternating Least Squares" (Sapper et al.).

The original comments by the reviewer are presented in red text, our responses are shown in black text, and significant additions to the manuscript are presented in blue text. References to line numbers below correspond to the manuscript version submitted August 2024 for review.

## 1 Reviewer 3 Comments

### 1.1

Since the article was submitted in 2022, the latest references are from 2021. I suggest that the authors perform a short review of some recent literature to incorporate some of the latest developments.

The following additions have been made to Section 1.1 to acknowledge recent literature in the field of positive/nonnegative matrix factorization (particularly for large datasets and with application to atmospheric composition analysis).

**Line 40:** "Traditional factor analysis methods are known to be computationally expensive. Steps to speed up factor analysis have been explored, such as randomization and the use of graphical processing units (GPUs) (Halko et al. (2011); Tan et al. (2018)). Developing efficient algorithms is especially critical in atmospheric mass spectrometry, as improvements in instrumentation and increases in the duration of their use in field campaigns has led to intractably large datasets. Currently, analysis of these datasets requires sacrificing data resolution or extensive manual preprocessing to operate within existing PMF software tools, and full analysis can routinely take days or weeks of computation time (Hopke et al. (2023)). As a result, a variety of approaches have emerged for efficient source apportionment of atmospheric mass spectrometry data. Algorithms to solve the nonconvex optimization posed by PMF range from gradient descent, block coordinate descent, and projected gradient methods (Guo et al. (2024)). Attempts at using supervised, ensemble machine learning approaches have been shown to be capable of replicating results from traditional (unsupervised) factorization methods while reducing computation time (Zhang et al. (2025)). Recently, Erichson et al. (2018) applied randomization to PMF and introduced a new method, randomized hierarchical alternating least squares (RHALS), to solve the unweighted PMF problem. In this paper, we test the application of RHALS to atmospheric concentration data that contain uncertainties. Accounting for these uncertainties as regression weights, we introduce a method of externally weighting and unweighting the data, which to our knowledge is novel in its application to RHALS. We consider the accuracy and the reduced computational costs compared to other PMF algorithms commonly used in the field of atmospheric science."

**Line 136:** "It is not feasible to span all possible variants that $\mathbf{T}$ can take. Thus, the problem is often simplified to considering only positive rotations (values of $\mathbf{T}$ greater than zero) and negative rotations (values of $\mathbf{T}$ less than zero). A rotational program in PMF2 called FPEAK uses the parameter $\phi$ to denote the rotation strength, with positive values leading to positive rotations in $\mathbf{W}$ (Paatero (1997)). Paatero further improved this method in the Multilinear Engine (ME) algorithm, where the strength of rotation is allowed to vary between factors (Paatero and Hopke (2009)). The pulling algorithm presented in Paatero and Hopke (2009) is a sophisticated rotational method; more rudimentary pulling methods that mimic varying the regularization of the factor matrices are presented in Paatero (1997) and Paatero et al. (2002). Recent attempts at controlling for rotational ambiguity have involved additional factorization of the time-series matrix $\mathbf{W}$ into a matrix incorporating shape regularization to reflect known diurnal patterns of factors and a diagonal scaling matrix (Nanra et al. (2024))."

### 1.2

Some aesthetic changes - Authors can work on improving the plots and figures presented in the paper and try to make them uniform across all the figures presented in the paper. Numbers and lines in some of plots are too small to be properly visible when printed out.

While the authors agree in spirit with the reviewer's comment, the untimely passing of the fist author, loss of access to his original figures, and the inherently stochastic nature of the methodology means that we are unable to reproduce and make stylistic improvements to the figures in the manuscript. However, all figures are vector graphics and can therefore be zoomed in on digital versions with high fidelity.

### 1.3

While not relevant for publication, the authors should consider making detailed documentation of the code written in MATLAB available for others to use with ease. This will enable a larger adoption of the proposed RHALS method.

The authors appreciate and agree with the suggestion. For the reasons mentioned in Section 1.2, we are unable to make changes to the GitHub repository containing the EW-RHALS MATLAB code. However, as the goal is to make EW-RHALS open source and widely available, we are currently porting the code from MATLAB to Python. At the time when we are ready to publish the Python code, we will ensure there is detailed documentation of the code and algorithms.

# 2 Topic Editor Comments

## 2.1

Please consider the few points raised in the last review. Regarding the figures, please also label all panels with (a),(b),...

All figures have been amended to include panel labels as needed.

# References


Erichson, N. B., Mendible, A., Wihlborn, S., and Kutz, J. N.: Randomized Nonnegative Matrix Factorization, Pattern Recognition Letters, 104, 1–7, https://doi.org/10.1016/j.patrec.2018.01.007, 2018.

Guo, Y.-T., Li, Q.-Q., and Liang, C.-S.: The rise of nonnegative matrix factorization: Algorithms and applications, Information Systems, 123, 102 379, https://doi.org/https://doi.org/10.1016/j.is.2024.102379, 2024.

Halko, N., Martinsson, P. G., and Tropp, J. A.: Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions, SIAM Review, 53, 217–288, https://doi.org/10.1137/090771806, 2011.

Hopke, P. K., Chen, Y., Rich, D. Q., Mooibroek, D., and Sofowote, U. M.: The application of positive matrix factorization with diagnostics to BIG DATA, Chemometrics and Intelligent Laboratory Systems, 240, 104 885, 2023.

Nanra, M., Saha, S., Shukla, A., Tripathi, S., and Kar, P.: Robust Shape-regularized Non-negative Matrix Factorization for Real-time Source Apportionment, pp. 192–201, https://doi.org/10.1145/3632410.3632457, 2024.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometrics and Intelligent Laboratory Systems, 37, 23–35, https://doi.org/10.1016/S0169-7439(96)00044-5, 1997.

Paatero, P. and Hopke, P. K.: Rotational tools for factor analytic models, Journal of Chemometrics, 23, 91–100, https://doi.org/10.1002/cem.1197, 2009.

Paatero, P., Hopke, P. K., Song, X.-H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, Chemometrics and Intelligent Laboratory Systems, 60, 253–264, https://doi.org/10.1016/S0169-7439(01)00200-3, 2002.

Tan, W., Chang, S., Fong, L., Li, C., Wang, Z., and Cao, L.: Matrix Factorization on GPUs with Memory Optimization and Approximate Computing, Proceedings of the 47th International Conference on Parallel Processing, 26, 1–10, https://doi.org/10.1145/3225058.3225096, 2018.

Zhang, Y., Fang, J., Meng, Q., Ge, X., Chebaicheb, H., Favez, O., and Petit, J.-E.: An Ensemble Machine Learning Approach for Predicting Sources of Organic Aerosols Measured by Aerosol Mass Spectrometry, ACS ES&T Air, https://doi.org/10.1021/acsestair.4c00262, 2025.